Collaboration based Function Prediction in Protein-Protein Interaction networks

Hossein Rahmani[1], Hendrik Blockeel[1,2], and Andreas Bender[3]

[1] Leiden Institute of Advanced Computer Science, Universiteit Leiden,
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
hrahmani@liacs.nl,blockeel@liacs.nl
[2] Department of Computer Science, Katholieke Universiteit Leuven,
Celestijnenlaan 200A, 3001 Leuven, Belgium
[3] Division of Medicinal Chemistry, Leiden/Amsterdam Center for Drug Research,
Leiden University
2333 CC Leiden The Netherlands
andreas.bender@pharma-it.net

1 Abstract

In recent years, much effort has been invested in the construction of protein-protein interaction (PPI) networks. One particular machine learning task that has been considered is predicting the functions of proteins in the network.

A variety of methods have been proposed for predicting the functions of proteins. A number of existing methods predict protein functions based on the assumption that interacting proteins have similar functions. Following this assumption, these methods use optimization criteria such as "minimizing the number of interacting pairs of proteins with no common function" in order to make predictions. For example, the Majority Rule considers this optimization criterion by assigning to a protein those functions that occur most frequently among its neighbors. Functional clustering methods first cluster the PPI networks and then assign the function shared by the majority of the cluster's proteins to all the proteins in the cluster. A common drawback of both approaches is that they rely solely on the assumption that the closer two proteins are in the PPI network, the more similar their functional annotation should be.

Our point of departure is that a biological process is the aggregation of each individual protein's functions, and the main idea behind our function prediction techniques is that topologically close proteins tend to have *collaborative* functions, not necessarily the same functions. We define collaborative functions as pairs of functions that frequently interface with each other in different interacting proteins.

We propose two methods that predict protein functions based on function collaboration. The first method calculates the collaboration value of two functions using a straightforward "Reward-Punish" strategy; the second method adopts a Self Organizing Map (SOM) for this purpose. We perform a comprehensive set of experiments that reveal a significant improvement (ranging from 3% to 17%) of F-measure values compared to existing methods.